The UAV Video Image Stitching Based on Improved Moravec Corner Matching Method

Chaokui Li[1,2], Gang Yang[1,2], Jun Wu[1,2] and Fang Wang[1,2]

[1]Center of Geospatial Information Science, Hunan University of science and technology, Xiangtan 411201, China

[2]Hunan State Engineering Laboratory of Geospatial Information，Xiangtan 411201, China

chkl_hn@163.com

Abstract- **As a kind of new image sources, UAV (Unmanned Aerial Vehicle) Video is used more and more widely. In this paper, the image matching based on corner feature is used to achieve image stitching. Moravec operator is a kind of simple and efficient method for corner feature extraction. In this paper we do special treatment on it, which covers the distributed conjugate points over full frame with virtual grid to improve the precision of corner feature extraction. In the process of image mosaic, we use fade in/out method to achieve video image seamless stitching.**

Keywords- UAV Video；Moravec；Feature Extraction；Image Stitching

I. INTRODUCTION

The UAV technology has developed rapidly since 1990s. UAVs are mainly designated to undertake highly specialized missions with the characteristics of being dull, dirty, and dangerous. It has undergone an explosion in military areas and attracted attentions of many civilian applications. For many civilian UAV projects, video camera is often used as image sensor to obtain clarity, details, and characteristics of ground surface features and provide observers with a real-time view of activity and terrain. Aerial video is quickly becoming an up-to-date source of imagery because of its low cost

In common with other image stitching, UAV video image stitching also needs to find out the overlap between adjacent frames, then complete image matching and mosaic, in which image matching is the key step. Over the years, many image matching algorithms were proposed, such as matching algorithm based on ratio, algorithm based on frequency domain correlation and feature-based matching algorithm, etc. Because of its strong robustness feature, the matching algorithm receives much attention. In this paper, we take corner-feature-based algorithm to implement image matching. Considering complexity, simple but effective Moravec operator is implemented in this paper to extract obvious point feature from left frame. First we introduce hierarchical matching based on pyramid into image matching process as the ambiguity solution. We do special treatment on this operator that covers the distributed conjugate points over full frame with virtual grid, to improve the precision of corner feature extraction. The lighting conditions outside for the UAV video data obtaining are basically the same. So we use fade-in/out to eliminate the traces of the overlap splice in the splicing process.

II. PROCESSING METHOD

Before video is in real use, one traditional pre-processing to video is re-sampling multiple frames from video stream at a fixed time interval to get appropriate connection points as an aim. If necessary, in a further video images processing of the previous removal of some" extra frame " or" harmful frame " is allowed.

The traditional method is video re-sampling technology which is based on inherent nature of video stream and based on time-stamped. In this project the highly automated algorithm is developed, which can produce the connection point and video sequence simultaneously. the algorithm differs from the video streams to the traditional pre-processing method, it through finding the right to complete the image matching algorithm for disparity vector and generates video images to determine the time marker, cleverly put the video image re-sampling and imaging matching processing are combined into a whole, the flow-chart of this algorithm is shown in Figure 1.

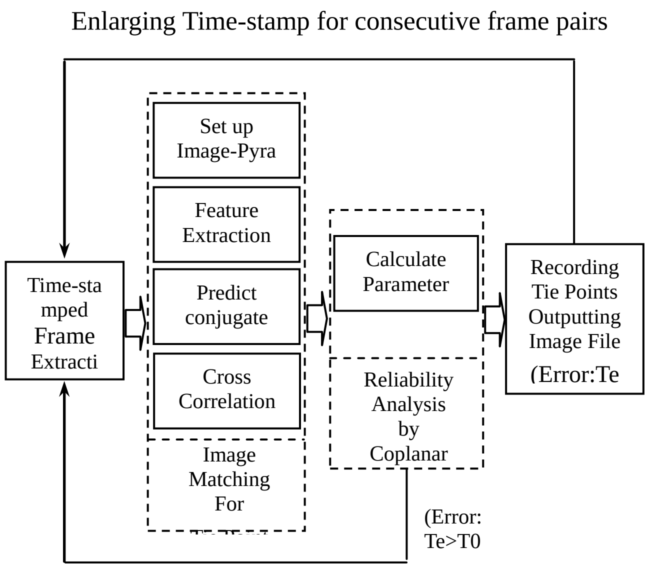

Fig. 1 Algorithm Steps to Generate Tie Points for Video Frames

A. *Performing Pyramid Transform on Key Video Frame*

One of the challenges for high-resolution UAV video is ambiguity solution which occurs in matching process due to repetitive patterns. Prediction and small search space around predicted location for conjugate point are effective strategies to ambiguity problems. In this paper we introduce “from coarse to fine” hierarchical matching based on pyramid into image matching process, in order to enlarge pull-in range as well as decrease sensitivity of the gray values to noise. A pyramid is a sequence of images of decreasing resolution which is repeatedly convolved an initial image with a set of low-pass filters “W” in term of digital signal processing. Many literatures have proposed different schemes to represent

and compute pyramid from original image, like Laplacian pyramids used to match scene images obtained under different illuminate conditions. Orientation energy pyramids are used to represent images with different scales and orientations. Extended spatio-temporal orientation pyramids are used to support the analysis of time varying imagery by defining the filtering over video volumes for added representational power. A simple method is recommended in this paper to represent the pyramid of the re-sampled video frames. Intensity value of each pixel in higher level pyramid is simply the intensity average of 3x3 image region in adjacent lower level pyramid, as shown in Fig.2. One of the reasons is that when we transform the adjacent points between pyramids, the location of adjacent points is still retrained because every pixel of the higher pyramid is exactly the centre covered with the 3*3 image region of the adjacent lower pyramid. Fig.3 shows the multi-resolution pyramid converted from a pair of source of images. Because of narrow-field-of-view aerial video, only three-level pyramid is implemented in this paper.

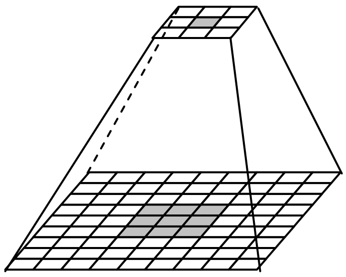

Fig. 2 Pyramid structure

Fig. 3 Performing pyramid transform on video frame

B. Feature Extraction

The prominent difference between different matching algorithms is probably the distinction between different matching primitives. This is because further matching technique like similarity measuring for primitive candidates, optimization for searching, etc., is usually based on selected matching primitives. The matching primitives selected are generally divided into two categories: either a window composed by gray value or the features extracted from each piece of image. The priority knowledge is often used in actual matching process. The algorithms above are called area-based matching (ABM) and feature-based matching (FBM) respectively. The algorithm used in this paper actually belongs to the former; however feature extraction is used in this paper to overcome weakness of the ABM algorithm. As we all know, although ABM has a high accuracy potential in well-textured image regions and in some cases the resulting accuracy can be quantified in terms of metric units, the weakness of ABM is the sensitivity of the gray value to changes in radiometry, the large search space for matching including various local extremes, and the large data volume which must be handled. So there may be some mistakes in block region and the poor result or the texture repeat. How to resolve the weakness is that ABM is extracted with two methods in this paper.

(1) Point-feature-extracting operator is firstly used in the left image (We call the first image left image and the second image right image for the adjacent frames re-sampled from the video), and only the window centered at those point feature extracted are considered for real ABM.

(2) The adjacent point of every extracted point feature is projected in the right image, which is based on hypothesized parallax vector. Only these local windows centered at these points can be considered for real ABM，which are in the searching region round with the adjacent points projected from extracted point features.

In recent years, all kinds of interest technologies are developed to extract point features from image, like Moravec, Hannah, Forstner, etc. Considering the complexity, the simple and efficient Moravec operator is used to extract obvious point features from left image. Steps include :

(1) Calculate the IV (Interest Value) for each pixel, namely, calculating the sum of the gray difference square along four different directions in the $w \times w$ image window (Fig. 4) centered with the pixel (c, r). The formula below shows the resulting V1, V2, V3, and V4, of which the minimum is taken as the IV for each pixel.

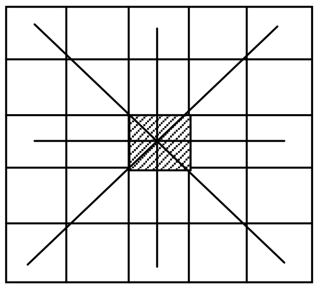

Fig. 4 Moravec operator

$$V_1 = \sum_{i=-k}^{k-1} (g_{c+i,r} - g_{c+i+1,r})^2$$

$$V_2 = \sum_{i=-k}^{k-1} (g_{c+i,r+i} - g_{c+i+1,r+i+1})^2 \qquad k = INT(W/2)$$

$$V_3 = \sum_{i=-k}^{k-1} (g_{c,r+i} - g_{c,r+i+1})^2$$

$$V_4 = \sum_{i=-k}^{k-1} (g_{c+i,r-i} - g_{c+i+1,r-i-1})^2$$

(2) Give a threshold M_T, and take the IVs these above M_T as point feature candidate. The principle of threshold selection is that the candidates must include feature points necessary not too much non-feature points.

(3) Determine point feature from candidates. Remove the points not the maximum in a certain search window, the pixel with the only point living is determined as final point feature.

We do a small but special treatment on Moravec algorithm in this paper, which covers the distributed conjugate points

over full frame with virtual grid (e.g. each cell of the grid is 40*40 pixels). Moravec operator is applied to each cell to extract the point feature if it exists. The following three reasons explain this step:

(1) Reduce the time costs for points feature extraction on numerous frames. Since our goal is to generate the tie points for the chains of frames, e.g. for three consecutive frames k_{i-1}, k_i, k_{i+1}, the tie points of the first stereo frame (k_{i-1}, k_i) are considered as the obvious point feature of the left frame of the next stereo frame (k_i, k_{i+1}), that is supposed as the k_i frame covered with virtual grid. The tie points of stereo frame is P= {p_i, i=1, 2, 3...n}, then Moverac operator is necessarily used in the image region of the grid cell with certain tie points. Only the "null" cell is involved in the point feature extraction. So the time cost is saved in point extraction of numerous images.

(2) Eliminate possible ambiguities. Once an amount of extracted point features are concentrated in local image region, the mixed texture, which easily causes ambiguity in further cross correlating and extrema locating, limits only one point feature to be extracted in each grid cell by using Moravec operator .

(3) Guarantee reliability analysis of matching result. The reliability analysis based on coplanar constraint is used for all the matched adjacent points to determine the generating accuracy of the tie points. To guarantee the coplanar constraint geometrically is robust and the adjacent points distributed over full frame averagely are very important. Fig. 5 shows the result when Moravec operator is applied to the frame over virtual grid.

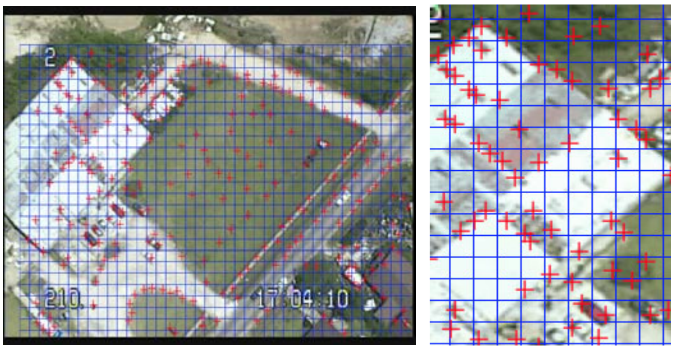

Fig. 5 Implementing Moravec operator to frame covered with virtual grid

C. Projecting the Adjacent Points

Base on hypothesized coarse parallax vector, we can predict the corresponding adjacent points in right frame for the extracted feature points in left frame. Fig. 6 shows the point feature extraction result in left image and the predicted adjacent points in right image. In this progress large data volume for ABM computation and search space for matching including various local extrema are reduced greatly and time is also saved. In addition, because extracted point feature is always distinct with respect to their neighborhood, stable with respect to noise, invariant with respect to geometric and radiometric influences ， so the sensitivity of gray values to changes in radiometry is greatly decreased.

D. Implementing Cross Correlation

Defining a match criterion plays an important role in each match algorithm. For ABM the similarity between gray value windows is usually defined as a function of the differences between the corresponding gray values. In this paper, the function is the cross correlation coefficient between the target window centered with the extracted point features and the match window centered with the points within the projected adjacent range. Its mathematical explanation is $\rho = S_{xy} / \sqrt{S_{xx} S_{yy}}$ (Where: S_{xy} is covariance function of target window and matching window; S_{xx}, S_{yy} is variance function of target or matching window respectively.)

$$\rho(c,r) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(g_{ij}g'_{i+r,j+c} - \frac{1}{mn}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}g_{ij}\right)\left(\sum_{i=1}^{m}\sum_{j=1}^{n}g'_{i+r,j+c}\right)\right)}{\sqrt{\left[\sum_{i=1}^{m}\sum_{j=1}^{n}g_{ij}^{2} - \frac{1}{mn}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}g_{ij}\right)^{2}\right]\left[\sum_{i=1}^{m}\sum_{j=1}^{n}g^{2}_{i+r,j+c} - \frac{1}{mn}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}g_{i+r,j+c}\right)^{2}\right]}}$$

Comparing the maximum value of cross correlation coefficient with the threshold, we then determine the adjacent point of each extracted point feature. Fig. 7 shows the corresponding result when cross correlation is implemented on Fig. 6.

Fig. 6 extracting feature and Predicting conjugate point

E. Image Mosaic

It is easy to form obvious seams at the boundary of the overlap region when we complete the image composition by directly using one of the two images in the overlap region. So the mosaic technology is necessary in image stitching. In this paper fade-in/out is recommended to add the pixel values of the overlap region by gradient coefficient. Take two stitched images I_1 (i, j) and I_2 (i, j) for example, the pixel values of overlap region I (i, j) can be expressed as:

$$I(i,j) = dI_1(i,j) + (1-d)I_2(i,j)$$

where, d is a gradient coefficient. The formula above shows that image transit from I_1 (i,j) to I_2 (i,j) with d changes from 1 to 0 gradually. So the stitch trace is eliminated to accomplish smooth transition between images.

Fig. 7 Tie point candidates from Cross Correlation

Fig. 8 Image stitching and mosaic

III. EXPERIMENTS AND RESULTS

According to the methods and procedure of video image processing introduced above, we used video image provided by United AOSI Company to do the experiment and we got the results as following. Fig.8 shows the satisfactory seamless stitching result of several video frames.

IV. CONCLUSIONS

With lower-cost UAV platform, cheap and high resolution images which focus on the area of interest and accommodate changing weather conditions can be obtained even when manned missions are not possible. In the processing of UAV video data, we use the developed Moravec operator and fade-in/out to implement image mosaic by carrying out the seamless stitching with high resolution. In this way we can accurately locate the place on fire and use the method as a good guide to any emergent rescue.

ACKNOWLEGEMENT

This paper was supported by Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (10R01 and 11I04).

REFERENCE

[1] Wang Han, Liu Zhi-gang. An automatic stitching method based on SIFT for UAV video image [J]. National Security Geophysics Series (IV) - geophysical environment detection and target information acquisition and processing, 2008:170-175.

[2] Zhang Jian-qing, Pan Li, Wang Shu-gen. Photogrammetry [M]. Wuhan University Press.

[3] H. Adelson , C. H. Anderson , J. R. Bergen . Pyramid methods in image processing [J]. RCA Engineer, 1984, 29-6:33-41.

[4] Peter J. Burt, Edward H. Adelson. The Laplacian Pyramid as a Compact Image Code[J]. IEEE Transactions on Communications, VOL. COM-3l, No. 4, April 1983:532-540.

[5] Gang Hong, Yun Zhang. Combination of feature-based and area-based image registration technique for high resolution remote sensing image [J]. Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International , 23-28 July 2007: 377 – 380.

[6] Moravec H P. Towars Automatic Visual Obstacle Avoidance[J], Int. Joint conf. of Artif. Intelligence, 1977: 584.

[7] Forstner. W. E. Gulch. A fast operator for detection and precise location of distinct points, corners and centers of circular features[C]. Interlaken: Switzerland Proceeding of Inter commission Workshop on Fast Processing of Photogrammetric Data, 1987.

[8] Gao Guang-zhu,Li Zhong-wu, Yu Li-fu, He Zhi-yong. Application of the Normalized Cross Correlation Coefficient in Image Sequence Object Detection [J]. COMPUTER ENGINEERING & SCIENCE, 2005; 27(3): 38—41.

[9] Wang Yong-min, Wang Gui-jin. Image Local Invariant Features and Descriptors. National Defense Industry Press, 2010.